This document is published in:

Graña, M. et al. (eds.) (2011) *Advances in Knowledge-Based and Intelligent Information and Engineering Systems*, (Frontiers in Artificial Intelligence and Applications, 243), IOS Press, pp. 1001- 1010.
Doi: http://dx.doi.org/10.3233/978-1-61499-105-2-1001

Group Behavior Recognition Issue, Feature Analysis on Defending Pick and Roll Basketball Move

Alberto Pozo<1>, Miguel A. Patricio<1>, Jesús García<1>, and José M. Molina<1> and Ignacio Refoyo<2>

[1] Group of Applied Artificial Intelligence,University Carlos III of Madrid, Colmenarejo, Spain
{alberto.pozo, jesus.garcia, miguelangel.patricio, josemanuel.molina}@uc3m.es http://www.giaa.inf.uc3m.es

[2] Sport and Performance Research Group, Faculty of Physical Education and Sport (UPM)

Abstract. The paper presents a group behavior representation and an application in the 2 vs. 2 basketball domain. Furthermore a set of forty features have been made from the raw information provided by the INEF12 Basketball Dataset. Moreover, from all these features we propose a selection using an algorithm to choose the best features to classify and predict the group behavior. The entire experimental test carried out with Hidden Markov Models algorithms could validate the proposed representation and features selection, in group behavior recognition and 2 vs. 2 basketball specific domain.

1 Introduction

Behavior recognition, with humans or with other kind of elements, is one of the most prolific fields of the current research. One of the typical restriction in these research is that must be only one element in the scene. The elements behavior is analyzed and the system recognizes the activity that the element is doing. There are a lot of papers with this objective, like [1] and [2]. But a lot of situations depend on more than one element; where elements have social behavior that depends on the others elements of the group; like team sports, animal social behavior (ethology), traffic analysis, etc. In all this case the individual behavior depends on the group behavior, so we need to study all the individuals together, we need to study them like a group.

There are a few works on this area; some of them manage the elements like a crowd; and try to recognize their behavior analyzing the shape, or some other fea-tures. In the other hand, some works manage each element like individuals that be-longs to the group; our proposal is in this way. In behavior recognition we could distinguish two different steps: features extrac-tion and features analysis. In the first one the system chooses and extracts the raw main information (from video cameras, GPS, RFID, TOF cameras, etc.). And in the second one the system uses this raw information to transform it and recognize the underlying behavior. This paper is focused on the second step, but also explains the extraction phase.

We propose a novel representation of the selected features that could be used to learn and predict the group behavior. Could be a lot of features in the scene, depend-ing on domain, and these features could be of several types, but in our opinion the location of the groups elements should be very meaningful in most

of the potential domains. All these features are structured in different levels, depending on if they are from an individual, from the scene or from the group. Our porpoise representation is structured in three different levels that are described below.
Some of the selected features are specific from the domain; but, in our opinion, features related with the location of the elements could be generalized for the most domains. To prove the useful of the representation, we have classified the group behavior of the dataset [13], with this objective we have used a well-known technique called Hid-den Markov Model [15]. So the representation could be useful for group behavior recognition issue. Also, there is a selection of the most representative features in the specific domain porpoise in [13].
The paper is organized as follows. Section 2 reviews related work. Section 3 de-scribes the problem issue. Section 4 shows the experimentation done. Conclusions and future work are drawn in sections 6 and 7.

2 Related works

Behavior recognition is one of the most prolific lines of research in the computer vision domain. There are lot of works that try to extract human behavior high level information from different types of sensors, especially from video [1] [2] and [8]; but also from other types of sensors like accelerometers, gyroscopes, etc. All this approaches are focused on individual human behaviors but humans are social beings, and there are lot of behaviors that depend on the other elements of the group. For this reason, these are a relative new area of research where the elements belong to a group. In these groups the behavior of each element depends on the be-havior of the rest elements, and there is a high level group behavior that arises from the combination of individual behaviors.
There are lot of situations where we could see this type of behavior like robocup in [16], team sports [6] [11] [14], military parade [5], automatic camera surveillance in public places like [3] [4] [17] etc. All these works could be classified in two main streams: logical reasoning and geometrical reasoning, depending on the type of fea-ture studied. Logical approaches like [7] and [4] are based on high level features that need more previous work; for example in [11] the players trajectory are discretized in main subareas, and there are some key movement like bloc that are detected. Then this high level information is analyzed to recognize the team behavior. These types of approaches use to be very effective but depending on high level features extraction system, and usually are too ad hoc solutions and the system could not be generalized. In the other hand there are geometrical approaches that use lower level features like raw position or some derivative feature like speed or acceleration. For example in [7] the authors use the players trajectories in an american football game to generate a discriminative temporal interaction matrix, using a 4-D tensor and a refactor 2-tensor kernel. This approach tries to classify five different team behaviors.
All these approaches are focused on the second step of the behavior recognition, where the features have been extracted yet, and the system uses these features to classify or predict the group behavior. So we need a first step level that extract from the raw data the selected features, for instance, extract from the video data the player trajectories in an american football match.
Another possibility is to use an existing dataset to develop, improve and test the approach. But in group behavior recognition there are not a lot of dataset that

could be adequate, so this is a very important difficulty that must be overcome. For example, papers like [7] use a non-public dataset called GaTech Football Play, and there are another commercial approach like Prozone, that have eight fixed cameras allocated in some soccer field that provide the trajectory of all players in the mach. Unfortunately we have not access to this information. In the other hand, the Namez Pers work [10] has three different parts related with three different sports: squash, basketball and european handball. The part related with the team sport European handball has the necessary information for the group behavior recognition research field. This information is composed by the one team players position and the team behavior in each team, for ten minutes video. Similar to [10] in [13] there are information of the player position in 2 vs. 2 basket-ball situation, information relative to the owner of the ball in each time and the type of attack performed (from eight different types). This dataset [13] was used in this paper to develop and test our approach. For the final classification and prediction we have used a well-known technique called Hidden Markov Models [15].

3 Group behavior recognition issue

Group behavior recognition is a novel field of research that comes from the elimina-tion of the one-element restriction in activity recognition issue. This field of research has a lot of potential domains such group sports, military intelligence, fauna behavior recognition, video surveillance, etc.
Behavior recognition research field have the characteristic that action happens in time. There is no isolated observation that must be classified, but a sequence of ob-servations where each one depends on the previous one. From this characteristic we could have two different problems: classification and segmentation. In the first one there are several segmented sequences, which each one have only one group behavior; in the other hand we could have only one sequence with many group behaviors, that must be segmented previously. This paper is focused on the first approach, where we have several segmented sequences.
Classification phase of the group behavior recognition is composed by two steps: extraction and recognition. In the first one the features of the system should be selected and extracted, and in the second one these features are used to recognize the behavior. The system could have a lot of types of features like position, individual action, trajectory, speed, color, etc. In this paper we are going to focus on the second step, there is only a short descrip-tion about how was the dataset construction process (feature extraction), and we try to show how this information could be used in the second step (behavior recognition).

3.1 A general problem description

Group behavior recognition could be applied in a variety of domains like is described above. All this domains have some common features describe bellow. There is a sce-ne S, composed by a number M of groups and a number F of features 1. These groups are performing groups behaviors from a set of behaviors bi called B 2. All these features and all behaviors are performed in time.

$$S_t = \{G_t^1, G_t^2, \ldots, G_t^M, F_t^1, F_t^2, \ldots, F_t^{O_S}\} | 1 \leq t \leq T \quad (1)$$

$$B = \{b_1, b_2, \ldots, B_p\} \tag{2}$$

Each group is composed by a number of elements Nm and a number of features OG that depends on the domain 3. Moreover, each group is performing a group behavior from B 2 in each time.

$$G_t = \{I_t^1, I_t^2, \ldots, I_t^{N_m}, H_t^1, H_t^2, \ldots, H_t^{O_G}\} | 1 \leq t \leq T \tag{3}$$

Finally, each element is composed by a number OI of features (also depending on the domain). All these features change in time, so in the instant t, the element n is defined by 4.

$$I_t^n = \{J_t^1, J_t^2, \ldots, J_t^{O_I}\} | 1 \leq t \leq T \tag{4}$$

So the scene has M groups and OS features, each of these groups are composed by NM elements and OG features, and each element has OI features. These features could be of four types: Positive, Boolean, Relative and Enum. Each type describes one type of elements feature:

- Positive, described by the equation 5; represent a feature that could be any value greater that zero. This typically could represent the positioning in one exe, some distance, etc.

$$\{p_{nf} | f = 1 \ldots F, n = 1 \ldots N, p = 0 \ldots \inf\} \tag{5}$$

- Boolean, described by the equation 6; could be any binary feature that represents if the element have or not some characteristic. For instance, if one player is the owner of the ball.

$$\{b_{nf} | f = 1 \ldots F, n = 1 \ldots N, b = \{0, 1\} \tag{6}$$

- Relative, described by the equation refeq:rnf; represent a feature which value must be between one minimum and one maximum value. For example, the luminosity re-ceives by a sensor.

$$\{r_{nf} | f = 1 \ldots F, n = 1 \ldots N, r = \{0, 100\} \tag{7}$$

- Enum, described by the equation refeq:enf; could be useful where the elements have some characteristic that must be from a list of values. For instance, the role played by a basketball player.

$$\begin{aligned} \{e_{nf} | f = 1 \ldots F, n = 1 \ldots N, e = \{1, 2, 3, \ldots, K\} | \\ K = \text{“}possible\, values\, of\, feature'' \end{aligned} \tag{8}$$

The features selected to be extracted is a very important decision that usually depends on the problem domain. However, there are some features that are noteworthy in most of the potential domains. For instance, the location of each element of the group is related with the group behavior in almost all cases.

Our approach could represent all the scene of the group behavior recognition issue, for classification and also for segmentation. In this paper, we have focused on the first one, where there is no only one sequence with several behaviors but several sequences with only one behavior. So in the learning phase, the system has lot of sequences S1, S2, etc. each one composed by a number of instants T1, T2, etc. and labeled with only one behavior bi from the set B 2. And in the predicting phase, one new sequence Sj is analyzed and his label bj is predicted.

3.2 Defending pick and roll move problem domain

This paper is focused on the 2 vs. 2 basketball domain, which is very appropriate because represent with few elements (only four) a complex domain, basketball match. Often, in team sports match, group behavior are performed only by a subgroup of the team, for example, in soccer, the defense line is composed by only four or five elements of the team, and that subgroup performs a lot of group behaviors that could be analyzed. Also, in basketball matches, one move called pick and roll could be played by only two players of the same team. In [13], this is the move that the defenders try to stop. So there are a sequence segmented in 23 different moves, each segment represents an instance of a five types of defense move. In these segments there are four players; two attackers and two defenders. The attacker couple tries to perform the same move (pick and roll) all time, and the defender couple tries to stop them with five different ways. These different ways are the group behavior that must be predicted, and are named: to fight/go over, to go below, push, change and trap.

3.3 Features description

Features extracted are the X and Y position and who is the owner of the ball in each time. X and Y are represented by positive feature type, see (5); and owners ball by enum features type, (8). With this raw information we have made forty features, composed by thirty nine positive features and one enum. These features describe the players positions (with eight features, four players and two axes); teams center posi-tions (with four features, two teams and two axes); all players center position (with two features); relative players positions with regard to the owners of the ball (with eight features); players velocity (with eight features, four players and two axes, and ten frames window); teams velocity (with four features, two teams and two axes, with ten frames window) and all players center velocity (with two features). At last, three features from the laplace invariant of the graph. Also one enum features indicates who is the carrier of the ball.
So in our proposal representation, there are M = 2 groups, N = 4 elements, each group has N1 = 2 and N2 = 2 elements respectively, the scene has five features OS = 8 (the carrier of the ball, the location and velocity of the center of all players, and the laplace invariant of the graph), each element has OI = 6 features (X Y positioning, X Y velocity, and X Y relative positioning) and each group have four features Og = 4 (X Y positioning and X Y velocity).
All the experimentation has been done with this dataset [13], but to do all the algorithms in the used framework Marlab, we have flattened all the representation in two dimensions like is described below. Time dimension has been leaved, and all others have been storage consecutively in the order describe in 9.

$$S_t = \{I_t^1, I_t^2, I_t^3, I_t^4, G_t^1, G_t^2, F_t\} \; with \; I_t^n = \{X_t^n, Y_t^n, XX_t^n, YY_t^n, Vx_t^n, Vy_t^n\},$$
$$G_t^m = \{X_t^m, Y_t^m, Vx_t^m, Vy_t^m\} \; and F_t = \{X_t, Y_t, Vx_t, Vy_t, H_t, L_t^1, L_t^2, L_t^3\} \quad (9)$$

Where X and Y represents the coordinates in the couch system; XX and YY repre-sents the coordinates of one elements with regard to the owner of the ball; Vx and Vy represent the velocity; H indicates the owner of the ball and L is the invariant laplace feature.

3.4 Features Selection

From the features provided from the dataset [13], we have made forty features. From all these features we have selected some of them following the algorithm described in [9]. In this algorithm there is a candidate features list, in this case compose by forty features describe above. Then the system calculates the accuracy using only one of the features, and selects the feature that has best outcome. This feature is delete from the candidate list. However, the system calculates the accuracy with the selected feature and each one of the candidates features, and again selects the feature that provides the best accuracy. The algorithm stops when the accuracy does not improve with any feature.

Fig. 1 shows the system results with only one feature, in the first step of the algorithm. Rows represent the sequences, and columns represent experimentation with one feature. Red square mean that the system did not classify the example right, and green square mean that the system classified correct the sequence. In

Fig. 1. One Feature Matrix

this case, the selected feature was number 22, which represents player four X loca-tion regard to the ball carrier. At the end of the features selection process, five fea-tures were selected. Fig. 2 shows the order of selected features and the system accura-cy.

4 Experimentation

For the experimentation we have used Matlab framework in which language we have the entire necessary algorithm to do the behavior recognition issue. First selected.png

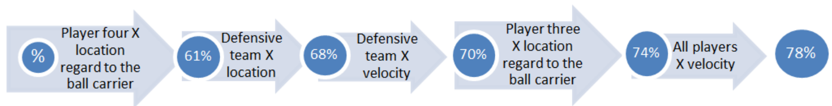

Fig. 2. Selected features

of all, the dataset has been loaded in to the system, with all the extracted features and data. INEF12Basketball dataset (which could be public downloaded from: http://www.giaa.inf.uc3m.es) provide 27 examples of a set of eight different group behaviors. For this experimentation, we have selected five of this behaviors per-formed by 23 of the examples. The rest of the behaviors have been eliminated because there are too few examples of them.

Then all the information is introduced into a matlab matrices, where are represented all the features in all time. So we have one matrix Mi (1 i 23) per each group action and one dimension vector with the behavior performed label. Each Mi matrix have FxTi dimensions, where F is the number of features selected and Ti is the dura-tion (in frames) of the behavior i. These matrices become the input for the describe algorithms, using leave-one-out cross validation [12].

It important to empathizes that not all the frames had been loaded, because only a few seconds per sequence have importance to the move. So per sequence, only one subsequence of few seconds selected by the expert had been loaded. For the experimentation, we have use well-known technique called Hidden Markov Model. For all experiment we have used a HMM with three Gaussians per class. Fig. 3 shows the accuracy of the system with one, two, three, four, five and six features, as we can see the sixth feature does not add more accuracy, so the system will use only four features. With this configuration we have 78% of accuracy, which rep-resent the number of example that the system classify in the right class. Blue bars represent the system accuracy with the best combination of the specific

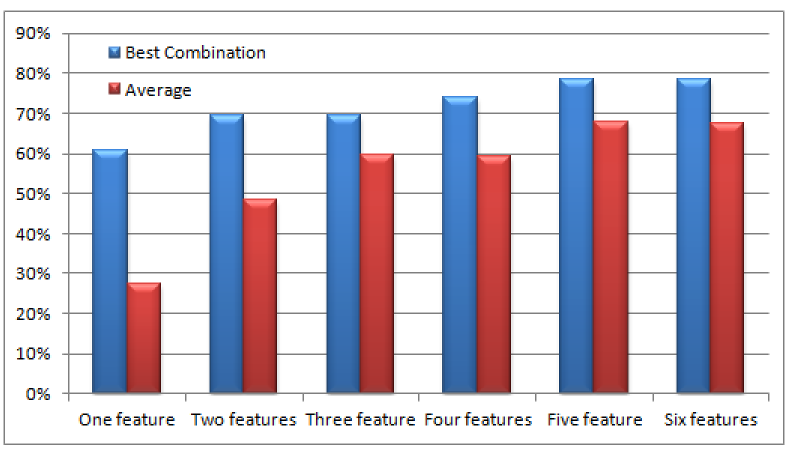

Fig. 3. System accuracy

number of features, and red bars represent the average accuracy from all possible combination of the specific number of features. As we can see, when there are five features selected, the system does not improve the accuracy adding one more feature, which causes that the features selection algorithm stops.

5 Conclusions

In this paper we have developed and test a system that uses some low level features (like positioning) to recognize high level features: group behavior. This approach is made on the 2 vs. 2 basketball domain, which have few elements in the scene however few features to analyze. Nevertheless, the approach could be applied to more complex system, and it could be effective in most of the potential domains. For instance, it is important to emphasize that in the basket domain, most of the group actions are performed by only a few players (not all the players must take part of all moves).

Also a features selection made could be useful in very different domains. Selecting the features could be very useful to improve the accuracy of the group behavior recognition system; if some features does not contribute any information, have too much noise, etc; Moreover, a simpler system (with less features) could be faster and more resistant system because the Hidden Markov Model could made the model with less gaussians.

6 Future work

All research approach have the target of generalize, in group behavior recognition there are two ways that this must be done: Domain and number of elements. The meaning of this is that all approach must try to be able of work in very different do-mains and with different number of elements in the scene. For this purpose, we think that INEF12Baskectabll Dataset proposed in [8] should be extended to more domains with different number of elements. In this way, all approach could be tested in some different domains, and we could see the scalability of the system. This is very important in group behavior recognition research field; because there are some authors that think some algorithms could not be applied in this issue for this reason [7].

Other important aspect that could be improved in this approach is to do it more solid. In could be achieved doing the system more resistant against tracker lost. For this purpose we need a system that could manage a variant number of elements in each group. This could be a very challenger and useful improvement.

In this paper, we have managed the problem of classification a segmented sequence, based on the group behavior. In our opinion, this is only the first step of one more ambitious issue. In the most case of the real world, we dont have the sequence segmented, so we a system that could be able to automatic segment (and classify) the group behavior. Obviously, the system must be trained with segmented examples, but then could be able to automatic segment a sequence with more than one group behaviors.

Acknowledgment

This work was supported in part by Projects CICYT TIN2011-28620-C02-01, CICYT TEC2011-28626-C02-02, CAMCONTEXTS (S2009/TIC-1485) and DPS2008-07029-C02-02.

References

1. J. Aggarwal and Q. Cai. Human motion analysis: A review. In *Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE*, pages 90–102. IEEE, 1997.
2. A. Bobick and A. Wilson. A state-based approach to the representation and recognition of gesture. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(12):1325–1337, 1997.
3. F. Cupillard, A. Avanzi, F. Bremond, and M. Thonnat. Video understanding for metro surveillance. In *Networking, Sensing and Control, 2004 IEEE International Conference on*, volume 1, pages 186–191. IEEE, 2004.
4. F. Cupillard, F. Bremond, and M. Thonnat. Behaviour recognition for individuals, groups of people and crowd. In *IEE Seminar Digests*, volume 7, 2003.
5. S. Khan and M. Shah. Detecting group activities using rigidity of formation. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 403–406. ACM, 2005.
6. R. Li and R. Chellappa. Recognizing offensive strategies from football videos. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 4585–4588. IEEE, 2010.
7. R. Li, R. Chellappa, and S. Zhou. Learning multi-modal densities on discriminative temporal interaction manifold for group activity recognition. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2450–2457. IEEE, 2009.
8. T. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 104(2):90–126, 2006.
9. M. Patricio and D. Maravall. Segmentation of text and graphics/images using the gray-level histogram fourier transform. *Advances in Pattern Recognition*, pages 757–766, 2000.
10. J. Pers. Cvbase 06 dataset: A dataset for development and testing of computer vision based methods in sport environments. 2006.
11. M. Perše, M. Kristan, S. Kovačič, G. Vučkovič, and J. Perš. A trajectory-based analysis of coordinated team activity in a basketball game. *Computer Vision and Image Understanding*, 113(5):612–621, 2009.
12. R. Picard and R. Cook. Cross-validation of regression models. *Journal of the American Statistical Association*, pages 575–583, 1984.
13. A. Pozo, J. García, and M. Patricio. Inef12basketball dataset and the group behavior recognition issue. *Distributed Computing and Artificial Intelligence*, pages 151–160, 2012.
14. A. Pozo, J. Gracía, M. Patricio, and J. Molina. A structured representation to the group behavior recognition issue. *User-Centric Technologies and Applications*, pages 47–57, 2011.

15. L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
16. F. Ramos and H. Ayanegui. Tracking behaviours of cooperative robots within multi-agent domains. In *AUTONOmOUS AGENTS*, page 45, 2010.
17. K. Schindler and L. Van Gool. Action snippets: How many frames does human action recognition require? In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.